BRIEF COMMUNICATION

Reliable change on the Dementia Rating Scale

OTTO PEDRAZA,[1] GLENN E. SMITH,[2] ROBERT J. IVNIK,[2] FLOYD B. WILLIS,[3]
TANIS J. FERMAN,[1] RONALD C. PETERSEN,[4] NEILL R. GRAFF-RADFORD,[5]
AND JOHN A. LUCAS[1]

[1]Department of Psychiatry and Psychology, Mayo Clinic, Jacksonville, Florida
[2]Department of Psychiatry and Psychology, Mayo Clinic, Rochester, Minnesota
[3]Department of Family Medicine, Mayo Clinic, Jacksonville, Florida
[4]Department of Neurology, Mayo Clinic, Rochester, Minnesota
[5]Department of Neurology, Mayo Clinic, Jacksonville, Florida

Abstract

A central role for neuropsychological evaluations is the measurement of change in cognitive functioning over time. However, change scores obtained from repeated neuropsychological assessments may be affected by normal variability because of measurement error and practice effects caused by repeated measurements. The current study uses reliable change estimates to establish normative rates of change on the Dementia Rating Scale from baseline to first follow-up testing among 1080 cognitively normal adults aged 65 and older. Results showed that a 6-point decline by European Americans or a 9-point decline by African American adults within a 9–15 month test-retest interval represents reliable change. Within a 16–24-month test-retest interval, a 7-point decline among European Americans or an 8-point decline among African American adults represents reliable change. In addition, preliminary cross-validation was performed in a clinical comparison sample of another 22 older adults. The findings are discussed in the context of potential clinical and research applications. (*JINS*, 2007, *13*, 716–720.)

Keywords: Cognitive aging, Alzheimer's disease, RCI, DRS, MOANS, MOAANS

INTRODUCTION

Establishing whether cognition changes over time is one of the primary roles of neuropsychological assessment. Repeated evaluations provide an index of cognitive stability or change following pharmacologic or behavioral treatments, surgical interventions for epilepsy or Parkinson's disease, resolution of toxic exposure or metabolic abnormalities, and cognitive rehabilitation or remediation after traumatic brain injury. In addition, repeated neuropsychological assessments can provide quantitative information on patterns of cognitive decline in neurodegenerative conditions, primary and secondary brain neoplastic processes, and cerebrovascular compromise.

Change scores obtained from repeated neuropsychological assessments, however, are a function of (a) the individual's true cognitive ability level at each testing point, (b) practice effects caused by repeated administrations of the same test, and (c) normal variability in test scores caused by measurement error. Each or all of these elements may vary according to additional factors such as individual characteristics of the test-taker, the duration of test-retest interval, and the disease process itself.

The guiding purpose for examining change scores is the identification of meaningful change over time. To validly attribute changes in test scores to an individual's true cognitive ability, it is necessary to account for potential practice effects and measurement error. To this end, a group of statistical procedures collectively known as reliable change indices (RCI) have been developed to characterize the expected distribution of change scores and to estimate the likelihood that an individual's score represents true change

(i.e., beyond chance fluctuations caused by the unreliability of the test).

The methods for optimally calculating reliable change have been debated extensively in this journal (e.g., Hinton-Bayre, 2000, 2004, 2005a; Maassen, 2004a; Maassen, 2004b; Temkin, 2004) and elsewhere (e.g., Hinton-Bayre, 2005b; Maassen, 2005; Maassen et al., 2006; McGlinchey & Jacobson, 1999). The reader is referred to these sources for a thorough discussion of the methodological and theoretical issues. Briefly, the debate has largely hinged on two issues: (a) whether modifications of the original Jacobson and Truax RCI method are preferable to a standardized regression-based approach (SRB), and (b) which standard error estimate is preferable for use as the denominator in the RCI methods. Regarding the former, Temkin et al. (1999) and Heaton et al. (2001) showed that the RCI method with correction for practice effects (RCIp) performed comparably to more complex regression-based methods, leading the authors to recommend the use of RCIp. Indeed, in a recent letter to this journal, one of the proponents of the SRB approach noted that the two methods are usually in agreement and most studies thus far have found minimal incremental accuracy outside of using the pretest score (Hinton-Bayre, 2005a). Given these observations, the RCIp method was deemed preferable for the current study.

The standard error of measurement of the difference (SE_{diff}) establishes a prediction interval around the test-retest change score when multiplied by a value from the z-distribution. As described by Iverson et al. (2003) and summarized by Maassen (2004a, 2005), whenever the post-test variance can be estimated and is not assumed to be equal to the pretest variance, the expression

$$SE_{diff} = [(SD_\times^2 + SD_y^2)(1 - r_{xy})]^{1/2} \qquad (1)$$

is preferable.

The goal of the current study is to provide normative rates of change from baseline to first follow-up testing on a commonly used measure of global cognitive functioning. The Dementia Rating Scale (DRS) is often used in clinical and research settings for early detection, differential diagnosis, and staging of dementia (Jurica et al., 2001). The DRS yields a maximum score of 144 points based on performance in five subtests. Mayo's Older Americans Normative Studies (MOANS) and Mayo's Older African American Normative Studies (MOAANS) have provided age- and education-corrected normative data for DRS scores (Lucas et al., 1998; Rilling et al., 2005); however, there are no reports in the literature regarding what magnitude of change constitutes a meaningful difference from a baseline level of performance. Heaton et al. (2001) noted that normative rates of change may not generalize from nonclinical to clinical samples, particularly if baseline test performance is dissimilar between groups. Therefore, this study also sought to evaluate the generalizability of the obtained reliable change values in a clinical comparison sample of individuals who were cognitively normal at baseline, but received a clinical diagnosis on follow-up assessment.

METHOD

Participants

Normal cognition

Participants with normal cognition included 1080 adults who took part in the MOANS and MOAANS series. Study criteria and recruitment protocols for these projects have been published previously (Ivnik et al., 1990; Lucas et al., 2005). Briefly, cognitively normal older adults were defined as community-dwelling, independently functioning individuals examined by their primary care physician within 1 year of study entry and who met the following criteria: (1) normal cognition based on self, informant, and physician reports; (2) capacity to independently perform activities of daily living based on informant report; (3) no active or uncontrolled CNS, systemic, or psychiatric condition that would adversely affect cognition, based on physician report; and (4) no use of psychoactive medications in amounts that would be expected to compromise cognition or for reasons indicating a primary neurologic or psychiatric illness. Substantial efforts have been made to follow MOANS and MOAANS participants longitudinally and only those who remained cognitively normal upon all follow-up assessments were included in the normal cognition sample. Every person was seen for at least one annual follow-up examination, although many had additional subsequent exams: 86.2% were seen a total of 3 times, 63.8% a total of 4 times, and 36.3% a total of 5 times.

Clinical comparison sample

Reliable change estimates derived from the normal sample were applied to a clinical comparison sample of 22 participants (21 European Americans, 1 African American) who were evaluated within an equivalent 9–24 month interval. The clinical comparison sample was 59.1% male and had a mean baseline age of 82.3 years ($SD = 6.3$, range: 70–96). All participants in this sample had normal cognition at baseline but received a clinical diagnosis upon their first follow-up assessment. Fifteen (68%) of them were subsequently diagnosed with mild cognitive impairment (MCI) or Alzheimer's disease (AD) based on consensus evaluation from a neurologist, family medicine physician, and neuropsychologist. Six (27%) participants had an interval stroke and one was diagnosed with Parkinson's disease between baseline and first follow-up. Clinical diagnosis was not assigned based on DRS scores.

Statistical Analyses

Within-group test-retest interval differences were evaluated using paired *t*-tests. Practice effects were defined as the mean difference between follow-up and baseline DRS

raw scores. Test-retest reliability (r_{xy}) was calculated as the correlation between baseline and follow-up DRS raw scores. Between-group comparisons were made using unpaired *t*-tests or χ^2 tests.

A reliable change index adjusted for practice effects (RCIp) was calculated with a 90% prediction interval (P.I.) based on the standard error as shown in Equation (1). The 90% P.I. was obtained by multiplying SE_{diff} by the corresponding value from the *z*-distribution (± 1.64). To adjust for practice effects, the resulting value was added to, or subtracted from, the mean difference between follow-up and baseline DRS raw scores. RCIp cutoff values were rounded to the next whole number for ease of interpretation and use. To further maximize its clinical utility, RCIp data are presented separately for participants who were retested within 9–15 months and for those retested within 16–24 months of their baseline evaluation. We chose to divide the sample in this fashion to provide clinicians with separate normative data for patients who return for follow-up testing close to one year after baseline, *versus* patients who return following a longer test-retest interval.

Recent population-based studies have found that older African American experience a more rapid rate of cognitive and functional decline than European Americans, even when sociodemographic factors are controlled statistically (Black & Rush, 2002; Moody-Ayers et al., 2005). In light of this literature, we derived reliable change values for the full normative sample and separately for European American and African American participants.

All data were obtained in full compliance with study protocols approved by the Mayo Clinic Institutional Review Board.

RESULTS

As shown in Table 1, European American participants were on average significantly older, more educated, and consti-

tuted a greater proportion of men. They were retested an average of 2 months later than their African American counterparts. There was no significant group difference in family history of dementia.

European American participants obtained significantly higher mean DRS scores than African Americans at baseline and at follow-up assessment. However, mean DRS practice effects were not significantly different between European American ($M = 1.3$, $SD = 4.7$) and African American participants ($M = 1.7$, $SD = 6.1$). Moreover, mean *within-group* DRS scores at each time point were in the average clinical range (scaled scores = 10) when using respective published norms. Mean DRS practice effects were not significantly different between men ($M = 1.3$, $SD = 5.2$) and women ($M = 1.4$, $SD = 4.5$), and were not associated with age ($r = -.01$, $p = .72$), years of education ($r = -.01$, $p = .87$), or family history of dementia ($r_{pb} = .05$, $p = .09$).

The mean DRS scores for the clinical comparison sample were 130.5 ($SD = 7.7$, range: 113–139) at baseline and 117.7 ($SD = 16.7$, range: 77–142) at follow-up testing. The mean difference in test-retest scores ($M = -12.7$, $SD = 13.8$) was significantly different ($t(21.1) = 4.77$, $p < .001$, adjusted for unequal variances) when compared to the mean test-retest difference obtained by participants with normal cognition ($M = 1.3$, $SD = 5.0$). When examining only those who obtained a follow-up diagnosis of MCI/AD, their mean test-retest difference ($M = -8.5$, $SD = 9.1$) was also significantly different ($t(14.12) = 4.21$, $p < .01$, adjusted for unequal variances) compared to participants with normal cognition.

Reliable change results are presented in Table 2. Among all participants, a decline of at least 7 points or an improvement of at least 10 points constituted reliable change at each test-retest interval. Within the 9–15 month test-retest interval, a decline of at least 6 points or an improvement of at least 9 points constituted reliable change for European American participants, whereas a decline of at least 9 points

Table 1. Demographic characteristics and Dementia Rating Scale (DRS) scores

	All participants ($n = 1080$)	European American ($n = 870$)	African American ($n = 210$)	Statistic[a]
Age at baseline (years)	75.6 (7.3)	77.2 (6.6)	68.7 (6.3)	16.94*
Sex (% male)	35.4	38.9	21.0	23.71*
Education (years)	13.2 (3.0)	13.4 (2.9)	12.4 (3.4)	4.70*
Family history of dementia (%)	32.8	34.0	27.2	2.91
Test-retest interval (months)	15.3 (3.0)	15.7 (2.9)	13.7 (2.5)	9.69*
DRS score (baseline)	135.5 (6.3)	136.6 (5.3)	130.9 (8.0)	9.84*
Range	99–144	110–144	99–144	
DRS score (follow-up)	136.8 (5.6)	137.9 (4.6)	132.5 (7.2)	10.26*
Range	104–144	113–144	104–144	

Data are shown as mean (*SD*) unless otherwise noted. Data on family history of dementia were missing for 24 participants and unknown for another 48 participants.
[a]*t*-test or χ^2 comparison of European American and African American participants.
*$p < .001$.

Table 2. Reliable change indices

	N	r_{xy}	SE_{diff}	90% PI	Practice effect	RCIp cutoff
Time interval: 9–15 months						
All participants	603	.71	4.9	±8.0	1.3	$-7 \geq RC \geq +10$
European Americans	425	.61	4.3	±7.1	1.2	$-6 \geq RC \geq +9$
African American	178	.70	6.0	±9.8	1.4	$-9 \geq RC \geq +12$
Family history of dementia						
Yes	175	.74	4.2	±7.0	.66	$-7 \geq RC \geq +8$
No	365	.73	4.7	±7.7	1.6	$-7 \geq RC \geq +10$
Time interval: 16–24 months						
All participants	477	.54	5.1	±8.3	1.5	$-7 \geq RC \geq +10$
European American	445	.51	5.0	±8.1	1.4	$-7 \geq RC \geq +10$
African American	32	.57	6.5	±10.1	2.8	$-8 \geq RC \geq +14$
Family history of dementia						
Yes	156	.53	5.0	±8.2	1.4	$-7 \geq RC \geq +10$
No	312	.55	5.0	±8.3	1.5	$-7 \geq RC \geq +10$

r_{xy} = test-retest reliability; SE_{diff} = standard error of the difference; 90% PI = 90% prediction interval; RCIp = reliable change index adjusted for practice.

or an improvement of at least 12 points constituted reliable change for African Americans. Within the 16–24 month test-retest interval, a decline of at least 7 points or an improvement of at least 10 points constituted reliable change for European Americans, whereas a decline of at least 8 points or an improvement of at least 14 points constituted reliable change for African Americans. Reliable change data were largely comparable between those with and without a family history of dementia.

As used earlier, RCI values define the boundaries of statistically significant change for individual test scores. RCI values can also be used to estimate the frequency of individual follow-up scores that should fall above or below the adjusted prediction interval. Thus, approximately 5% of follow-up scores are expected to fall above and 5% expected to fall below the 90% P.I. Across the 9–15 month test-retest interval, 22 (5.2%) cognitively normal European American adults declined by at least 6 points and 14 (3.3%) improved by 9 points or more. During the same time interval, 4 (2.2%) cognitively normal African American adults declined by at least 9 points and 11 (6.2%) improved by more than 12 points. Within the 16–24 month interval, 20 (4.5%) cognitively normal European American adults declined by at least 7 points, whereas 22 (4.9%) improved by more than 10 points. Finally, among cognitively normal African American adults, 1 (3.1%) declined by at least 8 points and 1 (3.1%) improved by more than 14 points.

In contrast, 16 out of 22 clinical comparison participants (72.7%) declined beyond the respective RCIp cutoff values. No one in the clinical comparison sample showed reliable improvement. When examining only those clinical participants who received a diagnosis of MCI/AD upon follow-up evaluation, 10 out of 15 (66.7%) showed reliable decline, whereas the remainder did not show reliable change.

DISCUSSION

This study sought to improve measurement precision when the Dementia Rating Scale is used to evaluate change in global cognitive functioning. There was no significant association between age, sex, years of education, and family history of dementia on DRS change scores. Among European American older adults, a 6-point decline in DRS scores over a 9–15 month period or a 7-point decline over a 16–24 month period represents statistically significant change. Among African American older adults, a 9-point decline over a 9–15 month period or an 8-point decline over a 16–24 month period represents statistically significant change. The subtle difference in RCIp values between European American and African American participants may be attributed to greater variability in DRS performance by African Americans, which in turn resulted in wider prediction intervals.

Few reliable change studies on normative samples have attempted to cross-validate the obtained results. Using the Wechsler Adult Intelligence Scale, Heaton et al. (2001) showed that RCI values obtained from a sample of 384 nonclinical adult participants did not generalize well to a new sample of 69 stable patients diagnosed with schizophrenia, resulting in larger-than-expected classification errors. The authors suggested that norms for change may generalize better if the two samples demonstrate similar baseline test performance. In an effort to provide preliminary cross-validation of the obtained DRS reliable change indices, we examined change scores in a sample of 22 participants who were initially classified as having normal cognition, but received a clinical diagnosis on follow-up consensus evaluation. In contrast to the proportion of normal cognition participants who demonstrated reliable decline (2.2% to 5.2%), approximately 73% of the clinical compar-

ison sample showed statistically significant decline, thus lending support to the generalizability of the current RCIp criteria for the DRS.

By defining the boundaries of normal chance fluctuation, we can identify with greater precision those individuals whose change scores may be worrisome for progressive cognitive decline. In the context of clinical evaluations, such information can guide test interpretation and diagnostic decisions. In the context of research evaluations, these findings may be used to establish trigger points whenever the DRS is used as a screening measure. For example, a reliable decline on the DRS could "trigger" a comprehensive assessment of that particular research participant.

It is important to note that the DRS reliable change estimates obtained in this study were calculated based on change scores from baseline to first follow-up visit. Thus, it is unclear whether these values will remain valid across various longitudinal time points. Nevertheless, these data may assist clinicians and researchers in determining whether an individual's change on the DRS from baseline to first follow-up represents a reliable indicator of cognitive decline or improvement.

ACKNOWLEDGMENTS

REFERENCES

Black, S.A. & Rush, R.D. (2002). Cognitive and functional decline in adults aged 75 and older. *Journal of the American Geriatrics Society*, *50*, 1978–1986.

Heaton, R.K., Temkin, N., Dikmen, S., Avitable, N., Taylor, M.J., Marcotte, T.D., & Grant, I. (2001). Detecting change: A comparison of three neuropsychological methods, using normal and clinical samples. *Archives of Clinical Neuropsychology*, *16*, 75–91.

Hinton-Bayre, A. (2000). Reliable change formula query. *Journal of the International Neuropsychological Society*, *6*, 362–363.

Hinton-Bayre, A. (2004). Holding out for a reliable change from confusion to a solution: A comment on Maassen's "The standard error in the Jacobson and Truax Reliable Change Index." *Journal of the International Neuropsychological Society*, *10*, 894–898.

Hinton-Bayre, A.D. (2005a). Methodology is more important than statistics when determining reliable change. *Journal of the International Neuropsychological Society*, *11*, 788–789.

Hinton-Bayre, A.D. (2005b). Commentary on Maassen's "Reliable change assessment in sport concussion research." *British Journal of Sports Medicine*, *39*, 488–489.

Iverson, G.L., Lovell, M.R., & Collins, M.W. (2003). Interpreting change on ImPACT following sport concussion. *The Clinical Neuropsychologist*, *17*, 460–467.

Ivnik, R.J., Malec, J.F., Tangalos, E.G., Petersen, R.C., Kokmen, E., & Kurland, L.T. (1990). The Auditory Verbal Learning Test (AVLT): Norms of ages 55 and older. *Psychological Assessment*, *2*, 304–312.

Jurica, P.J., Leitten, C.L., & Mattis, S. (2001). *Dementia Rating Scale-2: Professional Manual*. Lutz, FL: Psychological Assessment Resources.

Lucas, J.A., Ivnik, R.J., Smith, G.E., Bohac, D.L., Tangalos, E.G., Graff-Radford, N.R., & Petersen, R.C. (1998). Normative data for the Mattis Dementia Rating Scale. *Journal of Clinical and Experimental Neuropsychology*, *20*, 536–547.

Lucas, J.A., Ivnik, R.J., Smith, G.E., Ferman, T.J., Willis, F.B., Petersen, R.C., & Graff-Radford, N.R. (2005). Mayo's Older African American Normative Studies: Normative data for commonly used clinical neuropsychological measures. *The Clinical Neuropsychologist*, *19*, 162–183.

Maassen, G.H. (2004a). The standard error in the Jacobson and Truax Reliable Change Index: The classical approach to the assessment of reliable change. *Journal of the International Neuropsychological Society*, *10*, 888–893.

Maassen, G.H. (2004b). What do Temkin's simulations of reliable change tell us? *Journal of the International Neuropsychological Society*, *10*, 902–903.

Maassen, G.H. (2005). Reliable change assessment in sport concussion research: A comment on the proposal and reviews of Collie et al. *British Journal of Sports Medicine*, *39*, 483–488.

Maassen, G.H., Bossema, E.R., & Brand, N. (2006). Reliable change assessment with practice effects in sport concussion research: A comment on Hinton-Bayre. *British Journal of Sports Medicine*, *40*, 829–833.

McGlinchey, J.B. & Jacobson, N.S. (1999). Clinically significant but impractical? A response to Hageman and Arrindell. *Behaviour Research and Therapy*, *37*, 1211–1217.

Moody-Ayers, S.Y., Mehta, K.M., Lindquist, K., Sands, L., & Covinsky, K.E. (2005). Black–White disparities in functional decline in older persons: The role of cognitive function. *Journal of Gerontology: Medical Sciences*, *60A*, 933–939.

Rilling, L.M., Lucas, J.A., Ivnik, R.J., Smith, G.E., Willis, F.B., Ferman, T.J., Petersen, R.C., & Graff-Radford, N.R. (2005). Mayo's Older African American Normative Studies: Norms for the Mattis Dementia Rating Scale. *The Clinical Neuropsychologist*, *19*, 229–242.

Temkin, N.R. (2004). Standard error in the Jacobson and Truax Reliable Change Index: The "classical" approach leads to poor estimates. *Journal of the International Neuropsychological Society*, *10*, 899–901.

Temkin, N.R., Heaton, R.K., Grant, I., & Dikmen, S.S. (1999). Detecting significant change in neuropsychological test performance: A comparison of four models. *Journal of the International Neuropsychological Society*, *5*, 357–369.